Occasional Paper No 6 March 2014

Issues in Research Management and Administration


# Managing Open Research

Ray Kent
Liz Lyon
Mark Hahnel

Edited by Ray Kent, De Montfort University, Leicester

# Table of Contents



**About Issues in Research Management and Administration**
IRMA is a series of Occasional papers published by Association of Research Managers and Administrators (UK). For more information, please see www.arma.ac.uk

# Introduction

**Dr Ray Kent**
Director of Research, Business and Innovation,
De Montfort University, The Gateway, Leicester
LE1 5XY, United Kingdom
rkent@dmu.ac.uk

The management of research publications and research data has moved, over the past 5 years, from periphery to centre-stage in UK universities. Once considered to be the domain of librarians and a small but perfectly formed group of digital curation specialists, the topic can no longer be ignored by university managers. The clarion call from Government is this: embrace openness, publish freely (i.e. without pay-walls and passwords), and unleash your data! But are academics and their institutions listening to – and acting upon – this call to action? Speaking in December 2012 after a year of intense activity by the Department of Business, Innovation and Skills (BIS), aimed at ushering in the new era of 'open' research, Mr David Willetts MP, the UK Minister for Universities and Science, stated that open-access publishing was now 'sorted' from the perspective of Government. Next on the agenda would be the opening-up of research data. Unfortunately for Mr Willetts, it seems that not everyone was convinced of the wisdom of moving so hastily towards open access: academics sighed and got on with writing their papers (they were not to be rushed); the House of Commons grumbled (there were inconsistencies in the guidance provided to researchers); even the House of Lords took it upon itself to object (it didn't want to pass up the opportunity to pile in and criticise the Government); and thus the Minister's 2012 statement on open access being 'sorted' proved to be somewhat premature. As a result of this kerfuffle, Mr Willetts felt the need to engage further with the nay-sayers, whether in the national press[1] or in Parliament[2] ; and even to pop up in Berlin in November 2013 to make further pronouncements on the subject of open access[3] – albeit to an audience consisting entirely of converts to open access . Most recently, in his letter to Dame Janet Finch dated 23 January 2014, he seemed to think that the whole matter might finally be 'put to bed' with the assistance of Universities UK[4]. Only time will tell as to whether Mr Willetts is right.

Leaving Ministerial over-optimism on open access aside, it would appear that the era of 'big data' (i.e. large and/or complex data-sets) is now upon us, heralded by the Government's *Open Data Strategy 2012-14* (BIS, June 2012) and a landmark report entitled *Science as an Open Enterprise* (The Royal Society, June 2012). That report presented 10 recommendations for change, the clear intention of which is to facilitate – by legislation if necessary – a move away from a research culture in which data is viewed as a private preserve. Despite the rhetoric, such a move towards openness is unlikely to be quick or painless; researchers, whether in the public or private sector, have a tendency to guard their data jealously – or at least not to feel any great compulsion to publish it, which in the end will amount to much the same thing. Behind this lurks the fear of a loss of control over the use of the data, on the part of the author/originator, if text- and data-mining rights are granted freely or by licence from a publisher (for example, the recent proposal by Elsevier to allow text-mining of more than 11 million online research articles[5] ). This is not helped by the UK Government's stated intention, from April 2014, to make exempt from copyright the mining of texts for non-commercial purposes. Changes in the way in which we view research data are of course necessary if society is to exploit big data generated by researchers, as anticipated by the Open Data Strategy 2012-14. It appears we are entering a brave new world in which published data sets may assume an economic importance that was hitherto unimaginable. The question is: are we ready?

In this edition of ARMA's Occasional Papers Series entitled *Issues in Research Management and Administration*, we take a look at the management of open-access publications and open data from the perspective of different actors.

In the first article, **Ray Kent** (De Montfort University, Leicester) takes a look at how UK institutions are managing the transition to Gold open access, taking as an example his own university. He describes how the institution developed an open access policy and action plan, and rolled-out these documents to researchers, using a combination of briefings and online guidance. Ray explores how De Montfort University has utilised its block grant from RCUK, and how it plans to 'top-up' this grant. He also discusses how a university, with only modest resources, might create the necessary infrastructure with which to underpin Gold and Green open access; and lists some of the uncertainties (human and physical) that will need to be resolved in order to have a system that is fit for purpose.

In the second article, **Liz Lyon** (UKOLN and University of Bath) considers how UK universities have responded to the challenges associated with managing open research data. In many ways, these challenges resemble those relating to open-access publishing: the need to engage with a range of stakeholders involved in the 'lifecycle' of research data; the requirement to develop a coherent data policy and accompanying roadmap for effective management of data sets; the desire to create an appropriate infrastructure; and the need to ensure that



[1] See: http://www.theguardian.com/science/political-science/2013/apr/09/open-access-scientific-publishing-peer-review-scientific-publishing
[2] See: http://www.publications.parliament.uk/pa/cm201314/cmselect/cmbis/99/9902.htm
[3] See: https://www.gov.uk/government/speeches/open-access-research
[4] See: http://www.researchinfonet.org/wp-content/uploads/2013/02/BIS-Transparency-Letter-to-Janet-Finch-One-Year-On-Response-January-2014.pdf
[5] See: http://www.nature.com/news/elsevier-opens-its-papers-to-text-mining-1.14659

all of the activities are sustainable in an era of shrinking budgets. Liz also shows how such developments closely mirror those taking place in other countries, some of which are clearly ahead of the UK in this area.

The third article in our set comes from **Mark Hahnel** (figshare/Digital Science). Mark provides a critique of the state of play concerning the sharing of research data, which he couches in terms of a 'data problem' – or more precisely, a 'data sharing problem'. Mark argues that the current situation regarding research data is untenable, in that the lack of a consistent approach to data handling by academics, publishers and funding bodies leads to a substantial waste of scarce resources. He argues that the routine sharing of data by researchers will benefit those researchers, their peers in the scientific community, the funders of research, publishers and wider society. The key is to convince all of these parties that change brings with it significant advantages over the current system; in other words, that data sharing is a 'no brainer'.

I commend these articles to you, the reader, and hope that they will contribute in some small way to the national debates that are taking place in regard to managing open research.

### Acknowledgements

The authors wish to express their thanks to the referees for their assistance in reviewing the papers, and to Dr Emma Thompson (ARMA Executive Office) for her help in the production of this issue.

# Managing the transition to Open Access publishing: an institutional perspective

**Dr Ray Kent**
Director of Research, Business and Innovation,
De Montfort University, The Gateway, Leicester LE1 5XY,
United Kingdom
rkent@dmu.ac.uk

## Background and context

In the early years of the twenty-first century, prompted by developments in mainland Europe (notably the Budapest Open Access Initiative, 2002; the Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities, 2003) and the United States of America (the Bethesda Statement on Open Access Publishing, 2003), The Wellcome Trust and UK Government began to develop policies relating to open access publication of research results that they had sponsored in whole or in part. Prompted by The Wellcome Trust's statement of principles in November 2004, the first such policy from Government – issued by the Executive Group of Research Councils UK (RCUK) – appeared in June 2005, and dealt with the accessibility of research outputs, quality assurance, efficiency in terms of the use of public funds, and preservation issues. Following publication in June 2012 of a study entitled *'Accessibility, sustainability, excellence: how to expand access to research publications'* (the so-called 'Finch Report'), these policies, which started out as position statements and guidance to individual grant-holders, were given 'teeth' by a Government anxious to promote open access to university research funded by the tax-payer. In September 2012, a group of 30 research-intensive universities was allocated £10 million by the Department of Business, Innovation and Skills (BIS) in order to support the transition to open access publishing of journal papers and conference proceedings; the implication was that open access would, over a 5-year transition period, become mandatory for Research Council grant-holders that wished to publish the results of their research. This funding heralded other open research initiatives by BIS, notably the launch of the Open Data Institute in December 2012.

On 1 April 2013, block grants totalling £17 million were made available by RCUK to 107 universities and research institutes, as the first tranche of Government funding that is anticipated to be awarded each year until 2017/18. This latter funding was designed specifically to support the payment of article processing charges associated with 'Gold' open access[1] to journal articles. "To ensure ... administrative efficiency ... a cut-off point [was] set so that only institutions that are eligible for a block grant of £10,000 or more in [FY2017/18] will receive funding" (RCUK press announcement, 8 Nov 2012, paragraph 6).

Open access publishing in the UK was given a further boost in February 2013 when the Higher Education Funding Council for England (HEFCE), acting on behalf of the 4 UK higher education funding bodies, announced that it intended to consult formally in July 2013 on implementing an open access requirement as part of the post-2014 Research Excellence Framework (REF). The Funding Council went on to propose that for journal articles and conference proceedings to be eligible for assessment in the post-2014 REF, the post-print version would need to be made openly available – either via the Gold route or the Green route – immediately upon acceptance or publication. For the time being (i.e. until after REF 2020), research monographs are to be an exempted from these requirements, due to the limited progress made in this area towards achieving open access. The HEFCE consultation closed on 31 October 2013, and a final policy announcement is due in the Spring of 2014. This policy will affect papers published from 2016 onwards.

Despite this rapid progress, or perhaps because of it, these developments in open access have not been without their critics. The Government's unabashed promotion of Gold open access publishing, as exemplified by formal and informal guidance emanating from RCUK, has been severely criticized by both the House of Lords (February 2013) and the House of Commons (September 2013)[2]. However BIS has chosen to ignore these criticisms, and has continued with its advocacy of open access to both research publications and data[3]. What are university researchers and their institutions to make of all this? The responses are varied, ranging from the infectious enthusiasm of advocates and early adopters (as exemplified by the activities of OAIG, the Open Access Implementation Group[4]); to cautious optimism about the direction of travel (see, for example, ARMA's response to the HEFCE Consultation on Open Access in the post-2014 REF[5]); to weary resignation on the part of some academics – a bowing to the inevitable of 'yet another policy designed to constrain academic freedom'. Hence a survey of recent practice, entitled *'Acting and reacting … what are universities doing in response to RCUK's 2013 Open Access policies?'* (JISC, November 2013) suggested that universities were simply getting on with the job of putting in place the human and physical infrastructure that will be required if they are to become fully compliant with RCUK and HEFCE policies.

## Managing the move to Open Access publishing: the example of De Montfort University, Leicester

De Montfort University (DMU) researchers publish about 600 journal articles each year, of which it is estimated



[1] Gold Open Access, as defined by RCUK, refers to the open availability of a paper on the journal's website immediately upon the electronic publication of that paper. This access is made possible by the publisher, with or without the payment by the author of an article processing charge. To be compliant with RCUK policy, the paper must be amenable to interrogation by text-mining and data-mining tools, and allow unrestricted re-use of content provided that proper attribution is made. This contrasts with RCUK's definition of Green Open Access, whereby the post-print version of the paper (i.e. the final peer-reviewed version, as submitted to the publisher) is made available through an online repository without any restriction on non-commercial re-use, and within a defined period.
[2] See: http://www.publications.parliament.uk/pa/ld201213/ldselect/ldsctech/122/122.pdf and http://www.publications.parliament.uk/pa/cm201314/cmselect/cmbis/99/99.pdf, respectively.
[3] See, for example, the speech given by David Willetts MP at the Berlin Open Access Conference on 20 November 2013: https://www.gov.uk/government/speeches/open-access-research
[4] See: http://open-access.org.uk/
[5] See: https://www.arma.ac.uk/news/news04Nov2013

that 60 papers will acknowledge support from RCUK, of which roughly one-third will have a DMU researcher as the corresponding author. The University received no money from the BIS 'pump-priming' initiative in 2012, and its RCUK block grant allocations for 2013/14 and 2014/15 are vanishingly small at less than £20,000 per annum. This means that DMU has funding from RCUK that is sufficient to publish about 20 Gold open access papers in 2013/14. This number is rather modest when compared to the publishing ambitions of DMU's RCUK-sponsored researchers, as revealed by a web survey undertaken during July and August 2013.

As a result, the University has had to think creatively in terms of how it will proceed to manage the move to open access publishing (to put this into context, DMU's nearest neighbours – the University of Leicester and Loughborough University – will receive RCUK block grants of £160,000 and £176,000, respectively, in 2013/14, rising to £188,000 and £207,000, respectively, in 2014/15). In this instance, thinking creatively has meant seeking to place DMU at the heart of open access developments in the local area and nationally. In November 2012, the University's Directorate of Research, Business and Innovation – hereinafter 'the research office' – established MORe ('Managing Open Research'), a group comprised of research managers, library staff and IT services staff from the 3 Leicestershire universities. The aim of the MORe initiative is to allow the sharing of good practice, at local level, on open access publishing and open research data. This group met twice during 2012/13, with the focus of the first meeting being on preparing for Gold open access, and the second on promoting the wider open access agenda.

In January 2013, DMU, acting in partnership with Loughborough University, set up an electronic mailing list (MORE@jiscmail.ac.uk) devoted to much the same purpose as the MORe group, but 'meeting' remotely and thus without any geographical limitations to its discussions. The 2 initiatives have proven to be complementary to one another, with the small group setting being attractive to all, and the mailing list proving to be especially popular with UK research managers and administrators – whilst generating interest from as far afield as Australia. Membership of the MORE list is now fairly steady at about 150 members.

## Providing guidance to researchers at DMU

The initial challenge was to prepare a policy document and implementation plan for the University's Executive Board, in order to gain formal approval for activities that will ensure the effective management of open access journal articles and conference papers. Rather than re-invent the wheel, DMU's research office undertook a short email survey of other institutions' intentions, and used these to develop its own guidance for staff. In March 2013 the University adopted a simple policy and action plan that will provide an appropriate framework in which to operate during the period 2012/13 to 2017/18. The policy can be summarized thus: Green open access is to be the default route for both staff and Ph.D. students at DMU, with the mandatory deposit of the post-print version of each article in DORA (De Montfort Open Research Archive – i.e. the institutional repository) at the time of online publication, where possible; or following an embargo period, where imposed by the publisher. The action plan and timeline is shown in Table 1.

This process has proven to be straightforward; the greater challenge is to ensure that researchers are aware of the open access policy and action plan, and cognizant of the wealth of guidance documents or 'survival guides' (for example, decision trees for authors; information on the different types of copyright licence) that is now readily available on the internet. At DMU we have sought to accomplish this mainly by briefing sessions for authors run by the research office, with additional one-to-one support available through staff in the Library. These activities will be on-going throughout 2014, supplemented by guidance and reference materials on Gold and Green open access, placed on the Library's website.

## How does DMU allocate the funds provided by RCUK?

In keeping with RCUK's policy on the use of the block grant, DMU has set up an open access publications fund that is administered centrally by the Library on a 'first come, first served' basis. In order to gain access to the RCUK block grant, researchers at DMU must fulfil RCUK's eligibility criteria and should normally be the corresponding author on the paper in question; this latter requirement is also in place at the University of Leicester (Ian Rowlands, personal communication, 2013). Anecdotally, some other UK institutions appear to be more flexible than DMU and Leicester – particularly if the corresponding author resides overseas or does not have access to monies for the payment of APCs. It is unclear as to whether this situation will continue as open access becomes the norm.

De Montfort University intends to supplement the RCUK block grant from 2014/15 onwards by allocating to the central publications fund a portion of surplus income derived from research and knowledge exchange activities; a new policy on the distribution of surplus income will come into force in the Spring of 2014, and is

designed to complement the institutional policy on open access publishing. This additional funding stream will be utilized by RCUK-sponsored researchers if and when the block grant is exhausted, but crucially will also permit researchers that do not hold a grant from RCUK (or The Wellcome Trust) to publish their results in a journal that offers Gold open access.

### Building the infrastructure for Gold and Green open access

De Montfort University is currently setting up workflows for facilitating Gold open access publishing, that will likely make use of Converis, a current research information system or CRIS that will be available to DMU staff from the Summer of 2014. These developments will continue to be informed by Gold workflows emerging from institutions that took part in the JISC APC [article processing charge] pilot study during the latter part of 2013[6]. However, for the foreseeable future the University will manage the payment of APCs to publishers, rather than utilizing intermediaries for this purpose (NB. this is primarily for reasons of scale and not because DMU does not see any advantage in employing a third party to manage APCs on its behalf, at some point in the future).

The University intends to follow the RIOXX guidelines and metadata application profile[7], or a similar system such as that proposed by OpenAIRE[8], by recording in its CRIS key information such as: funding body; grant reference number; article-level metadata (e.g. language, publication date, format, publisher); researcher identifier(s) (e.g. ORCID); payment details for APCs, if any; type of copyright licence; and date of deposition in DORA, the institutional repository. Thus the information will be captured in such a way that it is capable of being exported in bulk to funders' databases such as the RCUK's Research Outcomes System, ResearchFish and Gateway to Research.

In terms of forecasting demand and budgeting for APCs, DMU has not entered into any pre-payment arrangements with publishers, nor other membership schemes designed to drive down the overall cost of Gold open access. However the University Library will monitor this area and keep open the option of entering into such arrangements where they can be justified in terms of scale and/or have the potential for major savings.

The workflow for Green open access is also being revised at DMU, made possible by an upgrade to DORA in 2013 and informed by developments outside the University, including emergent vocabularies and standards for open research. Regardless of whether a Gold open access option is available, authors will be required to deposit as full text their journal articles and conference papers in DORA and, if appropriate, may choose to deposit the full text in a subject-based repository. The Library will continue to verify that the correct version has been deposited in DORA, but responsibility for placing the correct version in a subject-based repository will remain with the author.

An important question, where an APC is paid, will be how to verify that immediate open access has been granted from the date of online publication. Other elements will also need checking: for example, has the journal published the acknowledgement of the funder's support; has a statement been included on access to the underpinning research data; has the correct type of copyright licence been used? All of these checks will need to be carried out manually, at least for the time being, and thus are highly resource-intensive.

This begs the bigger question of how much human and physical infrastructure (e.g. repositories, CRIS and other research management tools) is necessary and/or desirable in a medium-sized university, in order to oversee the transition to open access publishing, and thereafter to maintain the workflows and systems at an appropriate level? The human resources devoted to managing open access at DMU are certainly modest, consisting of 2 members of staff in the Library, 0.6 FTE in the research office, and varying numbers of staff from IT services, provided as required. All of these staff members have 'day jobs', for example oversight of the institutional repository, managing implementation of the CRIS and maintaining allied systems. This situation appears to be by no means unusual in the UK, except perhaps in the larger research-intensive universities, which tend to be more generously resourced in terms of central services. Is DMU's resource sufficient? The situation is being carefully monitored by the research office and Library to ensure that the University is compliant with funders' regulations, aware of the latest open access developments at national and international level, but also in touch with our researchers' current and future needs on open access. Only time will tell as to whether DMU has calculated correctly the resource requirements for the transition to full open access, but the University anticipates that further staffing will be required (most likely in the Library), as the system beds down.

### Conclusions

It will be apparent from the overview presented here that the management of open access publishing continues to be in a state of flux, both at national and local (institutional) level. This poses a considerable challenge



[6] See: http://www.jisc-collections.ac.uk/Jisc-APC-project/
[7] See: http://rioxx.net/
[8] OpenAIRE is the European Commission's portal for open access, see: https://www.openaire.eu/

to research managers and administrators, librarians and IT services staff responsible for facilitating the smooth transition to a predominantly open access publishing world.

Key to managing all of the above is the light-touch but effective monitoring of researchers' behaviour in relation to open access, which is not always straightforward to predict (e.g. periodic campaigns by activists to boycott certain publishers may result in some open access channels being closed temporarily or permanently to individuals or groups). Nor – given the rapidly changing landscape of open access – is this behaviour easy to record systematically; at best one might hope to capture the *zeitgeist*. Academics are also editors and reviewers, and in future one might expect them (perhaps aided and abetted by librarians, research managers and administrators!) to wield greater influence in terms of the open access policies adopted by journals.

De Montfort University has sought to be sensitive to the pressures faced by its academic staff, research staff and Ph.D. students when deciding what and when to publish; this remains a decision for individual researchers, albeit research administrators will need to be closely involved if the decision involves seeking access to the central publications fund or other internal resources.

The next few years will undoubtedly see the pace of change accelerating still further, as institutions attempt to bring together different elements of their research management systems to cope with the new demands of open research. Despite the undoubted difficulties in managing the transition to open access publishing there are many opportunities, not least for institutions to work collaboratively rather than in isolation. Grass-roots initiatives, such as MORe, show that this is already happening. For some universities this might eventually mean moving to shared services, for example at regional level, but first it will be necessary to ensure that all relevant internal systems are capable of 'talking' to one another – a not inconsiderable IT challenge in its own right, and one that will doubtless consume our energies for some years to come!

### Table 1: Managing Open-Access Journal and Conference Papers at DMU: Action Plan and Timeline, 2013-14

| Activity | Date | Ownership | Link to key elements of University OA policy (bold font), and intended outcome(s) |
|---|---|---|---|
| Provide staff and Ph.D. students with summary of DMU's approach to Open Access (link to profile-raising/training workshops - see below) | Jun-13 | Library and Research Office | Clarity regarding the University's stance on OA issues. |
| Establish a central contact point for OA queries: openaccess@dmu.ac.uk | Jun-13 | Library and Research Office | Single mailbox for all queries relating to OA, with monitoring by Library and Research Office staff. |
| Set up Central Publication Fund (using RCUK block grant) for RCUK-funded papers only | As soon as is practical | University Research Committee (URC); delegated to Library and Research Office to manage requests | Fund managed in line with RCUK expectations (details still to be determined). Open and transparent procees for applying for APCs (no peer judgement; 'first come, first served' basis until funds are exhausted). Staff are eligible to apply, either on their own behalf or on behalf of an RCUK-funded Ph.D. student. Quarterly reports generated for URC. |
| Manage requests for draw-down of article processing charges (APCs) from Central Publication Fund | As soon as is practical | Research Office, Principal Investigators, Co-Investigators | Provide support to Investigators to ensure they are acting in accordance with HEFCE, RCUK or other funder rules. Using Converis, monitor activity on ROS and equivalent systems used by other funders. Provide quarterly reports to URC, and feedback to inform RCUK review in third quarter 2014. |
| Monitor compliance and provide reports to URC. Give feedback to funders. | As soon as is practical | Library and Research Office | Clarity regarding the University's stance on OA issues. |
| Awareness-raising and training workshops on OA | On-going | Library and Research Office | Quarterly workshops held centrally, plus on-demand events in the Faculties. Target RCUK grant-holders in the first instance. |
| Establish and maintain web pages on OA, to include link to list of RCUK Gold OA-compliant journals (http://www.sherpa.ac.uk/fact), and a FAQs page | As soon as is practical | Library | Up-to-date advice for researchers at DMU. Accessible via Library's top-level page, Research Office's top-level page. |
| Agree a mandate for Green OA, i.e. full-text deposit of published manuscripts (post-print version) in DORA, the institutional repository | As soon as is practical | URC (Library to monitor) | Green OA is to be the 'default' route for staff and Ph.D. students at DMU. Mandate assures staff of continued freedom to choose where to publish (taking account of funders' rules), and complies with HEFCE rules for post-2014 REF. Reduces demand for APCs to meet Gold OA. |
| Enforce existing DORA mandate, i.e. peer-reviewed journal papers and conference proceedings to be deposited in DORA at time of publication | By 01-Jan-14 | URC (Library to monitor) | DORA mandate covers both staff and Ph.D. students. For the post-2014 REF, HEFCE will require capture of post-print version (final peer-reviewed version submitted to publisher) or publisher's version where permitted (NB. some papers will have embargoes). Move DORA towards higher % of full-text items. |
| Encourage deposit of back-catalogue of papers (as full-text items) | On-going | Library | Proportion of full-text items in DORA increases towards target of 50% (by 01-Jan-17). |
| Capture OA-related metadata about journal and conference publications produced by, or co-authored by DMU researchers | As soon as is practical | Research Office, Principal Investigators, Co-Investigators | Metadata available in DORA and Converis. Investigators supported to enter their outputs on ROS and equivalent systems. Minimum information required (to be agreed): funding body; grant reference number; OA status (Green/Gold); date of becoming a Gold OA publication, if appropriate. |

**Notes to Table 1:**
Green shading indicates action required in fulfilment of RCUK OA policy (March 2013) and/or draft HEFCE OA policy (February 2013).
DORA = De Montfort Open Research Archive (the institutional repository)

# Institutional responses to managing open research data

**Dr Liz Lyon**
Associate Director, Digital Curation Centre; Director UKOLN Informatics, University of Bath, Claverton Down, Bath, Somerset BA2 7AY, United Kingdom
e.j.lyon@bath.ac.uk

## Introduction

The growth of data-driven research exemplified by the UK Engineering and Physical Sciences Research Council EPSRC-funded eScience Programme, cyber-infrastructure developments in the United States and an increasingly "open" agenda from UK and US governments and European Directives, have combined to bring about changes in the research landscape in higher education institutions. This short paper provides an overview of some of the ways in which HEIs are responding to these changes and highlights some of the ongoing challenges in this area.

## External policy drivers

In the UK, we can trace early strategic reports which signalled the impending changes in research practice[1] and made recommendations[2] to universities and funding bodies. More recently in 2012, the influential Royal Society Report[3] made a series of recommendations to funders and institutions to promote open data. UK research funding agencies have collaborated to publish the RCUK Common Principles on Data Policy[4] and research councils and other funders have published their individual data policy directives e.g. Wellcome Trust. In the United States in May 2013, the Obama Administration published a ground-breaking Open Data Policy[5] which covers data and *"requires agencies to collect or create information in a way that supports downstream information processing and dissemination activities."* In June 2013, the G8 leaders signed the G8 Open Data Charter[6] which included five over-arching principles to support open data and innovation. All of these policy initiatives have greatly strengthened the local imperative to curate, manage and share research data as part of the wider open data environment, with significant implications for UK universities.

## Getting stakeholders engaged

Research-intensive universities are now actively responding; for those that receive direct grant funding from the Engineering and Physical Sciences Research Council (EPSRC), there is an explicit requirement to meet the nine expectations[7] set out by the funding council in their Policy Framework on Research Data. There are a wide range of stakeholders involved in the research data lifecycle with roles in the creation, collection, processing, analysis, curation and preservation of research data[8]. These include (but are not limited to) researchers, academic faculty and PIs, senior managers such as PVC/ Pro Rector / Vice-Provost Research, doctoral training centres, planning offices, research support staff, legal office, IT services, libraries and information services. There are also external stakeholders such as publishers, disciplinary data centres and learned societies who are part of the data lifecycle. Different institutions have approached the data stewardship challenge in contrasting ways. At some HEIs, a cross-institutional committee or working group has been created to oversee research data management planning, which can be an effective way to bring together key institutional players to progress implementation. At other institutions, university libraries have taken the lead in promoting research data management and in developing essential human and technical infrastructure support services. The engagement of senior managers e.g. PVC (Research) is key; getting their attention can be facilitated with a pitch focussed on collaborative data-driven research opportunities, compliance and reputational risks, and data quality and research integrity arguments.

## Developing a roadmap

One of the EPSRC expectations was for institutions to develop a Roadmap and the concept has been interpreted in different ways with exemplars collected by the Digital Curation Centre[9]. One such is from the University of Bath[10] which describes how the institution will respond and was approved by the Vice-Chancellors Group. Implementation of the Roadmap actions is central to the successful embedding of good research data management practice. This work was initiated by a Jisc-funded innovation programme project (Research360) and is now being taken forward by the University Library with additional institutional funding support for two new data support roles.

## Setting institutional data policy

One of the key elements of these Roadmaps is the development and adoption of an institutional data policy with accompanying guidance, which has been approved by the university research committee or appropriate group. Policies range from the aspirational to more pragmatic approaches "with teeth" and cover the roles and responsibilities of institutional stakeholders. Supporting procedures provide more detailed descriptions of the services and infrastructure which are required for managed access, data curation and stewardship. A selection of data policy exemplars[11] have also been collected by the DCC and include those from leading Australian universities such as Monash.



[1] Hey T. and Trefethen A. (2003). The Data Deluge: an eScience Perspective. Retrieved from: http://eprints.soton.ac.uk/257648/1/The_Data_Deluge.pdf
[2] Lyon E. (2007) Dealing with Data: Roles, Rights, Responsibilities and Relationships. Report for Jisc. Retrieved from: http://www.ukoln.ac.uk/ukoln/staff/e.j.lyon/publications.html#2007-06-19
[3] The Royal Society (2012). Science as an Open Enterprise. Retrieved from: http://royalsociety.org/policy/projects/science-public-enterprise/report/
[4] RCUK Common Principles on Data Policy, see: http://www.rcuk.ac.uk/research/Pages/DataPolicy.aspx
[5] http://www.whitehouse.gov/sites/default/files/omb/memoranda/2013/m-13-13.pdf
[6] G8 Open Data Charter, see: https://www.gov.uk/government/publications/open-data-charter
[7] EPSRC Expectations, see: http://www.epsrc.ac.uk/about/standards/researchdata/Pages/expectations.aspx
[8] Lyon E. (2012). Informatics Transform: re-engineering libraries for the data decade. International Journal of Digital Curation, 7 (1), pp. 126-138 http://www.ijdc.net/index.php/ijdc/article/view/210
[9] EPSRC Roadmaps, see: http://www.dcc.ac.uk/resources/policy-and-legal/epsrc-institutional-roadmaps
[10] University of Bath Roadmap for EPSRC, available at: http://www.bath.ac.uk/rdso/University-of-Bath-Roadmap-for-EPSRC.pdf
[11] Institutional data policies, available at: http://www.dcc.ac.uk/resources/policy-and-legal/institutional-data-policies/uk-institutional-data-policies

## Creating Data Management Plans

An increasing number of international research funders require a data management plan as a pre-requisite for a funding award e.g. National Science Foundation in the United States. The Digital Curation Centre has developed a setries of tools to help researchers to generate plans. There is a newly revised Checklist[12] which provides a series of questions and prompts to guide the researcher in thinking about their data. There is also a new (beta) version of the DMPOnline tool which provides templates to enable the researcher to draft a plan tailored to particular funder requirements[13]. The resulting plans will be of particular interest to research support offices which manage the grant funding process within institutions and employ CRIS systems for this purpose.

## Building data infrastructure

Whilst there is now a well-established group of open access repositories for research publications, there has been no equivalent infrastructure for data outputs generated from research. The Jisc Managing Research Data Programme has encouraged the development of data repositories and a number of institutions now have pilot repositories in place, based on varying repository platforms e.g. Southampton, Essex and Bath (ePrints), Lincoln (CKAN), Exeter (Symplectic), Oxford (DataBank). These repositories are at an early stage of implementation but there are some common infrastructure elements emerging which are essential building blocks to facilitate managed access to data, recognising that whilst many datasets may be openly available for wider use and re-use, other data should and must remain secure. Examples of the latter include data relating to human subjects e.g. patient data and data from selected longitudinal surveys in the social sciences.

The components of an effective and interoperable data infrastructure include using a common metadata schema to describe the datasets to enable discovery through data catalogues and registries; the adoption of common data file formats which may be discipline-specific e.g. the CIF in crystallography; the assignment of persistent identifers to datasets for publication and subsequent citation, such as Digital Object Identifiers or DOIs provided by the DataCite agency and author identifiers such as ORCID IDs. The variation in data-intensive capability of organisations or disciplines, can be assessed using the Community Capability Model developed by UKOLN Informatics and Microsoft Research[14].

## Capacity-building

Many institutions are finding that awareness of good data management practice is relatively low amongst the research community and that major advocacy programmes are needed. In paralell there is a requirement to develop these skills in new-entrant researchers and partnerships with doctoral training centres (DTCs) and research development teams, have begun to emerge. The Research360 Project at Bath built on the partnership with the DTC for Sustainable Chemical Technologies and developed a number of data training modules for researchers. Professional support staff also need to enhance their awareness and skills in the data space. The *Immersive*Informatics Project which is a collaboration between the University of Melbourne and the University of Bath, has led to the production of a modular course pilot where the programme includes two modules which are “immersive” with the support staff working alongside researchers, observing their day-to-day practice and helping them to curate and manage their datasets.

## Resourcing Data Management

The services and initiatives described above, clearly have significant resourcing implications for institutions at a challenging time of economic constraint. Some universities such as Bristol and Bath, have developed Business Cases to help to resource these data-centric activities, in terms of new roles and positions, new hardware and software platforms and the development and delivery of new services. Encouragingly, advertisements for novel roles such as Research Data Manager, Data Librarian and Data Scientist are appearing, suggesting that at least in some cases, these propositions have been successful. There are also opportunities for the development of shared data services across regional consortia. This model is being successfully developed in the Netherlands originally as 3TU.Datacentrum with the three technology universities of Delft, Eindhoven and Twente , and recently expanding to include the DANS data centre, to create a national service called Research Data Netherlands.

## Conclusions

We are at an exciting point in managing the development of open data services. There is an interesting coming together of top-down policy drivers from funding agencies and government with bottom-up tools and code from researchers and developers at the coal-face. The challenge for institutions and for global organisations such as the Research Data Alliance, will be to join up these initiatives into a cohesive and co-ordinated whole, to ensure that the research opportunities and new knowledge capital from open data, will be fully realised.



[12] DCC (2013). Checklist for a Data Management Plan, v.4.0. Edinburgh: Digital Curation Centre. See: http://www.dcc.ac.uk/resources/data-management-plans
[13] DCC (2013). DMPOnline tool. See: https://dmponline.dcc.ac.uk/
[14] Community Capability Model for Data-intensive Research. See: http://communitymodel.sharepoint.com/Pages/default.aspx

# The Status of Data in Academic Research

**Dr Mark Hahnel**
Digital Science, The Macmillan Building,
4 Crinan Street, London N1 9XW, United Kingdom
m.hahnel@digital-science.com

## Introduction

There are very few, if any, discoveries each year in academia that come about without building on concepts and ideas that have been previously published in academic journals. This is the natural progression of research. However, this is often limited to building on top of conclusions or ideas, as opposed to the actual research itself. Current dissemination of research is largely based on making available pdf-based summaries of key findings, as opposed to the actual research outputs and raw data behind the graphs. In order to track a diverse array of academic outputs, they must persist on the Internet. One way to do this is via the minting of Digital Object Identifiers (DOIs) by trusted repositories. These managed links overcome the problem of 'link rot', which has been shown to occur at c. 10%/year for non-traditional outputs[1]. This article addresses the current problems created by a lack of data sharing in academia. We also look at the incentives structure and potential solutions for improving the quality of academic outputs across all fields of research.

## The State of Data

Academia as a whole accepts the fact that there is a 'data problem'. Research that has been published in academic journals, invariably as text and image-based summaries, creates multiple concerns across all domains.

For example:

**i. Reproducibility**
Of critical importance is the ability of other research groups to reproduce the published findings. This cannot be achieved if all of the information is not available. This information can include raw data sets, precise versions of the software used and correct metadata for each file. This problem was recently highlighted when Begley *et al*. (2012) tried to understand why clinical trials for cancer therapies had the highest failure rate compared to other therapeutic areas. They tried to reproduce landmark studies and found that scientific findings were confirmed in only 11% of cases[2].

**ii. Improving the power of data-sets**
Current research-publishing techniques do not support the ability for subsequent researchers to make new analyses of the same data. This includes combining the data with other data sets, and for uses that may not have been anticipated by the original producer or collector. Recent moves by publishers, such as the launch of *F1000 Research* (Science Navigation Group) and *Scientific Data* (Nature Publishing Group), may go some way towards facilitating improvements in this area by encouraging the presentation of data-sets as a central part of an article.

**iii. Negative data**
Perhaps the most depressing part of academia is the waste of research outputs. So much funding and researcher time goes into doing experiments that produce null results. However null results aren't a bad thing: the researchers are conducting their experiments on the basis of good prior knowledge of the field and a well-formulated hypothesis founded on this. The problem here is that no single academic can be right all of the time. So when they carry out experiments (often at great cost, both financially and in terms of time) that do not confirm their hypothesis, where does this research go? The simple answer is nowhere. 'Null' or 'negative data' generally goes unpublished. There are some efforts in this space that have been largely unsuccessful[3]. It can be postulated that the reason for this is a lack of incentives. This suggests that the impact system and associated reward is wrongly configured. This is a self-perpetuating problem: if the researchers are basing their hypotheses on the published literature, they may be wasting time and money, building upon research that can be false positives. This will inevitably produce further null data.

**iv. Clinical trials**
The problem with *not* publishing negative data is that it has a significant impact in clinical trials. Failure to do so has led to drugs with placebo-, or even detrimental effects being released to the market, at great profit. Whilst there has been progress in the timely publication of clinical trials, fewer than half of the clinical trials funded by the US National Institutes of Health (NIH) were published in a peer-reviewed journal within 30 months of trial completion, and a third remained unpublished after 51 months. Additionally, only 22% of trials had released mandatory trial summary results on ClinicalTrials.gov within 1 year of completion of the trial[4].

**v. Detecting fraud**
Data forensics can identify where data sets have been doctored or fabricated. By using algorithms that look at the likelihood that a set of results occurred naturally, outliers, or data-sets that seem unnaturally perfect, can be interrogated. This could actually be applied to the data sets that have been made available in the published



[1] Hahnel, M. (2013). Referencing: the reuse factor. Nature, 502, 298–298. doi:10.1038/502298a
[2] Begley, C.G. and Ellis, L.M. (2012). Drug development: raise standards for preclinical cancer research. Nature, 483, 531-533. doi:10.1038/483531a
[3] O'Hara, B. (2011). Negative results are published. Nature. doi:10.1038/471448e
[4] Ross, J.S., Tse, T., Zarin, D.A., Xu, H., Zhou, L. and Krumholz, H.M. (2012). Publication of NIH-funded trials registered in ClinicalTrials.gov: cross-sectional analysis. BMJ, 344, d7292. doi:10.1136/bmj.d7292

literature. However, retractions of papers based on these techniques would go some way in dissuading academics from making available their raw data[5].

**vi. Plagiarism**
Use of another academic's data without giving proper attribution is very difficult to detect if the raw data aren't available. The main reason why none of the above has been addressed previously has been a lack of resources and technology to support the release of research data in the format it was generated. Since the birth of the Web and the Cloud, this is no longer true: among 3,247 scientists surveyed anonymously in the United States, 0.3% admitted to falsifying data and 1.4% admitted to plagiarism[6].

**vii. Animal experiments**
An area that is often overlooked is the impact that this 'non-sharing' mentality is having on research involving animals. While animals involved in trials are kept in the most humane conditions possible, the loss of animals in research that produced null or un-reproducible data is something that can be improved upon. Making all of the research outputs (and notably, the data) publicly available at some point will massively improve the efficiency of research in this area.

## The path to a solution?

The way to address the 'data problem' can be separated into bottom-up and top-down approaches; or incentives and requirements, carrots and sticks. The incentives for making available research data already exist for the most part, or are emerging with new technology. The problem then becomes one of raising awareness amongst the research community. There is a need for academic institutions and funding bodies to make sure that all of the academics for whom they are responsible, know about the following:

- Sharing detailed research data is associated with an increase in citation rate[7]
- It is possible to track a diverse range of metrics demonstrating impact at many levels[8]
- The intensity of data-set reuse has been steadily increasing since 2003[9]
- The technology exists to allow researchers to make all of their research outputs available openly online, in a time-efficient manner
- The academic reward system is changing[10].

The question of what constitutes data is a grey area. Non-traditional research outputs such as videos are seen as the raw data for many researchers. Scholars are increasingly sharing their raw data and not traditional research outputs through repositories such as figshare, Dryad and GenBank, each of which allow for the citation of data-sets, videos, genetic sequences and other files that traditional publishers can struggle to accommodate. This suggests that the incentivised, bottom-up approach is having some success; but it is generally agreed that in order to make widespread change, there needs to be requirements at the funder level, ensuring that researchers are disseminating the content in a responsible and fair manner.

The push-back from researchers, not wanting to make available the data behind their publication, has led to funders introducing mandates in order to ensure that the research they fund has as much impact as possible. Data-management plans for research, detailing what data will be created, and outlining plans for data sharing and preservation, are now a core requirement of grant applications for a long list of funding agencies around the globe. These include the US National Science Foundation, the NIH, NASA, the UK Biotechnology and Biological Sciences Research Council, the UK Medical Research Council and The Wellcome Trust.

The Force 11 Amsterdam Manifesto on Data Citation Principles (2013) states that: "A data citation in a publication should resemble a bibliographic citation and be located in the publication's reference list"[11]. This has long been accepted as the standard for treating non-traditional research outputs in the same way as traditional articles[12]. However, a look at the most recent citations of figshare data (Table 1) shows that this recommendation is not well enforced by publishers or authors, with only 25% citing the data in the reference list.

As the number of research outputs that are made available online increases, the flood of information requires more filters to allow researchers to find the content they need in an easy manner. The previous measure for this is citations. This obviously is still a very important metric at a paper level, but it should not be the only measure. However, to date, citations remain the most important currency for the progress of an academic's career. The impact factor is just one way of tracking academic impact. While it is generally accepted that getting published in a journal with a high impact factor is a sign of quality, it does not ensure that the work will have an impact. The reverse of this is also true. Publications in journals with no impact factor – and even no peer review – have had huge impacts. The *h*-index has been adopted by Google Scholar as a way



[5] Stevenson, M., Mostertz, W., Acharya, C., Kim, W., Walters, K., Barry, W., … and Potti, A. (2009). Characterizing the Clinical Relevance of an Embryonic Stem Cell Phenotype in Lung Adenocarcinoma. Clinical Cancer Research: An Official Journal of the American Association for Cancer Research, 15, 7553–7561. doi:10.1158/1078-0432.CCR-09-1939
[6] Martinson, B.C., Anderson, M.S. and de Vries, R. (2005). Scientists behaving badly. Nature, 435, 737–738. doi:10.1038/435737a
[7] Piwowar, H. A., Day, R.S. and Fridsma, D.B. (2007). Sharing detailed research data is associated with increased citation rate. PloS One, 2, e308. doi:10.1371/journal.pone.0000308
[8] Adie, E. and Roe, W. (2013): Enriching scholarly content with article-level discussion and metrics. figshare. http://dx.doi.org/10.6084/m9.figshare.105851
[9] Piwowar, H.A. and Vision, T.J. (2013). Data reuse and the open data citation advantage. PeerJ, 1, e175. doi:10.7717/peerj.175
[10] Mervis, J. (2010). NSF to Ask Every Grant Applicant for Data Management Plan. Science Insider, 1–2. Retrieved from: http://www.citeulike.org/group/13989/article/7847447
[11] For further details see: http://www.force11.org/AmsterdamManifesto
[12] Altman, M. and King, G. (2007). A Proposed Standard for the Scholarly Citation of Quantitative Data. DLib Magazine, 13, 1082–9873. doi:10.1045/march2007-altman

Table 1: Location of data citations in traditional scholarly publications

| Paper | Journal | Location of citation to figshare object |
|---|---|---|
| http://arxiv.org/pdf/1312.0910.pdf | arXiv | Availability Section |
| DOI: 10.1371/journal.pcbi.1003399 | PLOS Computational Biology | Code and data |
| http://arxiv.org/pdf/1311.4764.pdf | arXiv | Data Accessibility Section |
| DOI: 10.1111/geb.12132 | Global Ecology and Biogeography | Data Accessibility Section |
| DOI: 10.1111/1365-2745.12198 | Journal of Ecology | Data Accessibility Section |
| DOI: 10.1111/1365-2435.12174 | Functional Ecology | Data Accessibility Section |
| DOI: 10.1167/13.5.3 | Journal of Vision | Data Accessibility Section |
| DOI: 10.1371/currents.out-breaks.264e737b489bef383fbcba-ba60daf928 | PLOS Currents | Data Accessibility Section |
| DOI: 10.7717/peerj.138 | PeerJ | Data Deposit Section |
| DOI: 10.7717/peerj.215 | PeerJ | Data Deposit Section |
| DOI: 10.7717/peerj.148 | PeerJ | Data Deposit Section |
| DOI: 10.1098/rspb.2013.0554 | Proceedings B | Discussion |
| DOI: 10.1021/jo401316a | The Journal of Organic Chemistry | Experiment Section |
| DOI: 10.1371/journal.pone.0069741 | PLOS ONE | Genetic Analyses Section |
| DOI: 10.1128/genomeA.00172-13 | Genome Announcements | Genome announcement Section |
| DOI: 10.1128/genomeA.00325-13 | Genome Announcements | Genome announcement Section |
| DOI: 10.1128/genomeA.00953-13 | Genome Announcements | Main body of text |
| DOI: 10.1111/2041-210X.12089 | Methods in Ecology and Evolution | Materials and Methods |
| DOI: 10.1371/journal.pone.0061937 | PLOS ONE | Methods |
| DOI: 10.1371/journal.pone.0074770 | PLOS ONE | Methods Section |
| DOI: 10.1371/journal.pcbi.1003381 | PLOS Computational Biology | Methods Section |
| DOI: 10.12688/f1000research.2-262.v1 | F1000 Research | Methods Section |
| DOI: 10.1016/j.ibmb.2013.11.002 | Insect Biochemistry and Molecular Biology | Methods Section |
| DOI: 10.3897/JHR.33.5204 | Journal of Hymenoptera Research | Methods Section |
| DOI: 10.1111/2041-210X.12117 | Methods in Ecology and Evolution | Methods Section |
| DOI: 10.1021/sb400066m | ACS Synthetic Biology | Reference List |
| http://arxiv.org/abs/1311.3523v2 | arχiv | Reference List |
| DOI: 10.1045/november2013-vierkant | DLib Magazine | Reference List |
| DOI: 10.1139/cjfas-2012-0372 | Canadian Journal of Fisheries and Aquatic Sciences | Reference List |
| DOI: 10.1017/S0031182013001121 | Parasitology | Reference List |

| Paper | Journal | Location of citation to figshare object |
|---|---|---|
| http://arxiv.org/abs/1311.3523v2 | arχiv | Availability Section |
| DOI: 10.1045/november2013-vierkant | DLib Magazine | Reference List |
| DOI: 10.1139/cjfas-2012-0372 | Canadian Journal of Fisheries and Aquatic Sciences | Reference List |
| DOI: 10.1017/S0031182013001121 | Parasitology | Reference List |
| DOI: 10.1155/2013/508965 | ISRN Renewable Energy | Reference List |
| DOI: 10.1016/j.procs.2013.05.451 | Procedia Computer Science | Reference List |
| DOI: 10.1371/journal.pone.0067460 | PLOS ONE | Reference List |
| DOI: 10.7287/peerj.preprints.54v1 | PeerJ | Reference List |
| DOI: 10.1177/1464419313492317 | Journal of Multi-body Dynamics | Reference List |
| DOI: 10.1007/978-3-642-25643-1 | Springer Earth System Sciences | Results Section |
| DOI: 10.1371/journal.pone.0068337 | PLOS ONE | Results Section |
| http://arxiv.org/pdf/1308.6056.pdf | arχiv | Results Section |
| http://arxiv.org/pdf/1308.6320.pdf | arχiv | Summary and future work |
| DOI: 10.1186/1471-2164-14-797 | BMC Genomics | Supporting Data |

of measuring an individual's academic impact. It does not take into consideration the journal impact factor, or where the articles were published. Interestingly, when extrapolated to a journal level, Google Scholar's rankings demonstrate that several publications with a high impact factor also rank highly, based on their *h*-index. However, journals with low or no impact factor, such as *PLOS One* or the *arχiv*, may perform equally well[13].

This is just one example that demonstrates that multiple metrics can provide a more detailed view of the impact landscape. This reasoning has led to an increase in the use of metrics that seek to measure the impact of academic outputs, known as 'altmetrics'. Altmetrics are a measure of the volume of the conversation on the Web, in relation to research. This can come in the form of social media mentions, or news articles. This impact is measured in real time, giving it a distinct advantage over citation counts, which have a 2 year lag-time before any filtering based on impact can be applied.

## Conclusions

With the growth in data generated by more complicated technology, and larger or more complex experiments, researchers must come up with new ways in which to analyse the research. The tools for analysis are becoming key products of the research and can be re-used in the same way that data can. These tools are increasingly available in the form of scripts and software. The recent growth of computer code as an important research output has raised many questions about how to give credit to the academics that created it. In order for researchers to truly exploit their research budgets and the outputs they generate from these budgets, technology needs to continue to evolve, to make dissemination of research outputs as easy as possible. By mandating that all publicly-funded research is made openly available, the practice of data sharing should become integral to the scholarly workflow. If this is the case, the power in linked open research data should provide evidence to private research bodies that open data can have enormous economic, as well as societal benefit.



[13] See: http://googlescholar.blogspot.co.uk/2013/07/2013-scholar-metrics-released.html

ARMA (UK)
Sheraton House
Castle Park
Cambridge CB3 0AX

T +44 (0)1223 370052
E enquiries@arma.ac.uk
www.arma.ac.uk